A Complex Network Perspective on the World Science System

Scott W. Cunningham and Jan H. Kwakkel[1]

Abstract: This paper discusses capabilities for a systematic overview of world science delivered from the use of new output indicators of science and technology. The data may be usefully structured using a complex network perspective on national publication and international collaboration. This paper uses a random sample of publication data from 2009 to provide a timely update on world activities in science. A mixed predictive and descriptive approach is used in analyzing the data. A variety of methods including structural network analysis, and network regression, are used in the exploration of this sample. Insights are gained into key participants in world science, their positioning in a network of collaborative relationships, and the resultant morphology of the network which emerges from a mixture of random and geographic factors.

Keywords: international collaboration; complex networks; scientific publication; national indicators; world science

1 Introduction and Research Questions

The increasing proliferation of data online and on the internet is giving us an unprecedented overview of world activities in science and technology. Relevant data is being collected online from publication databases, but also emails, internet pages and online repositories of models and scientific metadata. World activities have expanded dramatically in scope, with an unprecedented rise in the number of countries and regions which are actively taking a part in scientific research. Research has yet to grasp the systematic production of science and technology, yet the comprehensive scope of science and technology output from publication to patent to product may make it possible to provide a comprehensive survey of the science system.

Such data provides a potentially extraordinary insight for science policy and national foresight activities. Finally, there are a number of items of folk wisdom which warrant further investigation. It is widely believed that scientific collaboration is a good thing, and yet there is little understanding of how to stimulate collaboration, and why international collaboration has expanded at such a dramatic rate. New findings are increasingly being sought from multidisciplinary research, but the measurement of multidisciplinary research is still relatively new. National governments invest in strategic technologies, such as nanotechnology, yet the consequences and impacts of these strategic impacts along a production chain of knowledge remain unclear.

Increasingly, researchers are adopting a complex networks perspective to understand and interpret these scientific outputs. Investigations into networks across a range of domains reveal how complex behaviors emerge out of the interaction of simple parts. Networks are composed of nodes, as well as

[1] Faculty of Technology, Policy and Management; Delft University of Technology; 2600 GA Delft; The Netherlands

edges. In the context of scientific networks, nodes are repositories of knowledge at the level of a country, or more disaggregate level such as region, organization or researcher. Edges are the known or inferred collaborative activities between parts of the science system. A complex network perspective reveals repeating structures across multiple scales of time as well as space. The complex network perspective also confronts the tension between apparent global structure, and the randomness of small-scale and local processes in scientific discovery. Even the smallest discoveries can play a critical role in a cascading production of new knowledge.

The scanning of millions of scientific abstracts is now routine; further, some databases of science now recognize over 200 distinct fields of science. Processing and distilling this information necessitates the development of new data collection scripts, and new software and algorithms for data analysis. Efforts in data collection can be usefully coupled with new visualization capabilities developed for displaying complex networks. Likewise, new application programming interfaces afford the use of geographic data, and the display of world networks of scientific activity.

The goal of the paper is to provide new insights into the current configuration of the world science system, using a complex network perspective. The paper is organized by first considering previous research efforts in the area of world science systems. The paper proceeds with the discussion of the data used in the analysis. The paper offers a mixed perspective, offering both descriptive results, as well as an effort to develop and test hypotheses concerning the world network. The specific findings are divided into a discussion of activities of the nodes of the world network; here countries and sovereign states. Complementary results are provided concerning the edges and links of the network; here the collaborative relationships between countries. The paper conclude with ideas for future research, including open questions concerning the world configuration of science. The paper also offers some observations concerning international collaboration in science, and the support of collaboration through policy initiatives.

2 Previous Work

Many previous researchers have used a complex network perspective to analyze science and technology. The most fundamental manifestation of complex networks are power law distributions. A number of well-known bibliometric distributions take the form of a power law. These distributions include the Zipf law [1] (determining the frequency of words), the Bradford law [2] (determining the utilization of research), Lotka's law (determining the productivity of individual researchers) [3] and the Matthew effect (determining citations earned per publication) [4]. In related work, de Sola Price demonstrated preferential attachment in scientific collaboration, presaging many modern investigations into random graphs [5]. Scientific collaboration itself has been a vehicle for constructing data sets of complex networks, as evidenced for instance by Newman [6]. Complex networks have also cast insight into the structure of knowledge in science, as well as the evolution of international collaboration in scientific activities. These items are discussed more fully below.

The initial impetus for national studies of science and technology was the desire to create a national accounting system, whereby public funding for scientific investment could be wisely and transparently recorded [7]. A vision of a comprehensive science advisory system which predated modern computers was offered by Vannevar Bush, a U.S. science advisor [8, 9]. Additional early efforts in scientific reporting of national outputs began with UNESCO. UNESCO consultants contributed proposals for developing standardized and unified measures of scientific inputs and outputs [10]. Such outputs are still used today

in the Frascati manual. Member states of the OECD broadly followed suit [11]. Science indicator systems have been developed by a number of nations including Australia [12, 13], Canada [11], Malaysia [14], the Netherlands [15], the United Kingdom [16], and the United States [17]. Brazil is developing a data infrastructure which appears to be a model for tracking individual level contributions [18, 19]. There are also equivalent indicators for the European Community of nations [20].

National investigations have readily been extended to surveys of world science as well. Research at the country level has demonstrated some of the strongest world collaborative links. Early investigations into the world science system are offered by Francis Narin on behalf of the National Science Indicators program of the United States National Science Foundation [18]. Other researchers have explicitly examined the world context of research and development in the emerging and reforming economies, including Brazil, India, Russia, China and South Korea [21, 22].

Prior work has also discussed the meaning and measurement of collaboration and knowledge exchange itself. Cautions in using publication databases have been raised. Particular attention has been given to the multi-scaled nature of collaboration: cooperation can occur both within and between levels of activity. Significant efforts have been made to develop better schemes to fractionate and appropriate credit publication activities across partners. The partial and incomplete nature of publication as a comprehensive measure of collaboration has been investigated [23]. The substantial and often inaccessible gray literature revealing informal ties between organizations has been discussed [24]. Bozeman discusses the challenges of transferring technology between organizations, noting the role of distinctive characteristics in technology producers as well as consumers [25]. Dasgupta and David explicitly discuss the role of information and incentives in facilitating successful exchanges of knowledge [26]. Despite concerns about measuring collaboration, many researchers have nonetheless acknowledged the fundamental role that scientific collaboration between organizations plays in the larger system of science and technology. Lewison has been concerned with the impact of science funding on the production of new knowledge [27]. Buisseret et al. discuss how and whether science policy facilitates additive new inputs to knowledge which would not already occur [28].

International collaboration in science has strong implications for the design of national policy. Further, the explosion of international collaboration beginning in the 1990s has also received attention, with researchers attempting to explicate reasons for the sudden growth in international collaborative activity. Leydesdorf argues that the burst in international activity is the result of self-scaling activities on a vast, international network [29].

Independently of national and international inventories, some researchers have examined the organization of knowledge in world science. Leydesdorf and Rafaol produced a network map of the major disciplines and fields of science, showing the systematic relationships between subject categories [30]. Visualization efforts at Sandia national laboratories also provide a useful and encyclopedic understanding of the relationship between fields and subfields in science [31, 32]. Other researchers have mapped knowledge using citations between related bodies of knowledge. Henry Small and others have studied research frontiers, deriving these bodies of fundamental knowledge from network structures derived from inter-paper citation patterns [33, 34]. Van Raan explored the fractal dimension of citation and information spaces [35].

3 Data

The goal of this paper is to update previous work on world scientific output using a relatively recent sample of world science from 2009. The paper embraces a complexity perspective, building on previous work which confronts the world science system as a vast network. The goals of this paper are mixed. Some of the outputs are solely descriptive: the authors believe simple descriptive or reporting measures of leading nodes and edges in the world scientific network have considerable significance for formulating policy. However we also offer a more predictive perspective on the data by advancing and testing hypotheses concerning the structure and emergence of the world network.

This investigation uses publication data from the Web of Science. The data is based on a random sample of forty thousand publications selected from 2009. The sampling procedure first develops a sample using a comprehensive title selection strategy. Then, out of this comprehensive pool of all 2009 articles, selected abstracts are downloaded using a unique abstract identifier. This sampling procedure is relatively free from bias. Nonetheless, there are few potential threats to validity which should be considered. The most general issue concerns the degree to which the World of Science database is a representative sample of world science as a whole. A second more narrow concern is the prospective updating scheme for 2009 articles. At the time of writing this article in 2011, articles from 2009 are still being indexed and abstracted. The update schema is uncertain; the desire to include relatively recent articles may introduce a systematic bias in the sampling. The third and final potential concern is the database partitioning used by Thomson Reuters, the provider of the Web of Science data. Thomson Reuters divides the database into virtual partitions involving 100 thousand articles. Queries which exceed this length are automatically truncated. The nature of this virtual partition is not apparent to users, and therefore there is a potential threat of a systematic bias in article sampling.

Collaboration occurs at multiple organizational scales [23]. This survey of the world scientific network attempts to examine collaboration at the broadest, most aggregate level of analysis. We therefore consider the nation as the unit of scientific performance. This choice of aggregation does not deny the utility of also examining regional, organizational, departmental, or individual levels of analysis. Ideally, science policy might construct a comprehensive system of indicators which relates macro-scientific activity to the choices and outputs of individual researchers. This however is a topic for further extended discussion in the literature.

Analysts using the Web of Science have traditionally faced difficulties in making a complete and comprehensive attribution of papers to higher levels of organization. Earlier records in the Web of Science included only the organizational addresses selected by the first author. More recently, the Web of Science included a complete list of authors, as well as a complete list of organizations contributing to the paper. Unfortunately a unique mapping from author to organization is still not provided. Most recently, starting in 2008, the Web of Science provides a unique mapping of author to organization. More than 82% of the papers provide this complete, unambiguous mapping. It appears that in 2010 and beyond, this mapping will be uniformly provided for all articles. The complex network results discussed in the following paper builds upon this newer, higher quality data concerning organizational attribution.

Papers are often written across organizations or even countries. The following procedures are used for crediting the papers across organizations. Papers are first credited equally across contributing authors. If a given author belongs to multiple organizations, the authorship still counts only once. The author's share is

divided equally among organizational addresses. Each organizational address may then be uniquely mapped to a country. This crediting procedure will often result in several countries receiving partial credit for a paper.

In the following sections we report key descriptive and predictive findings on the world network, based on the sample of 40 thousand articles. Findings are divided into two sections: nodes and edges. Findings about nodes are typically presented in list format. Findings about edges are typically presented in matrix format. Lists and matrices as outputs from technology mining activities have been previously discussed [36]. Reporting of nodes occurs at two levels – national, as well as at the organizational level of funding agencies. Edges involve national collaboration, but also linkages between national science and funding agency. Edges are also potentially multi-valued, since collaborations may be distinguished by the type of knowledge they produce. We use a number of different techniques to analyze and present this data, including simple tabulation, network regression, structural network analyses, and network visualization.

4 Results from Nodes: National Performance and Funding

We first develop an aggregation scheme for the nodes of the graph. The United Kingdom, a country comprised of countries, is aggregated from its constituent countries – England, Northern Ireland, Scotland, and Wales. The European Union is also aggregated from its constituent countries, including the United Kingdom. Taiwan and the People's Republic of China are aggregated to China.

The first finding concerns the major publishing countries in the sample. We provide several aggregate groupings to the data. For instance, we group the member countries of the United Kingdom, and further group the United Kingdom within the European Community as a whole. The Chinas are also reported separately and as an aggregate unit. The results are given in Table 1.

Table 1. Total Publication Output by Country

Rank	Country or Community or Grouping	Total Publication	Percentage
1	European Union	10494.1	32.0%
	Germany	1926.9	5.9%
	United Kingdom	1883.9	5.7%
	France	1305.8	4.0%
	Italy	1157.8	3.5%
	Spain	982.6	3.0%
	Netherlands	571.3	1.7%
2	USA	8432.1	25.7%
3	Chinas	3357.8	10.2%
	Peoples R China	2842.5	8.7%
	Taiwan	515.3	1.6%

4	Japan	2038.3	6.2%
5	Canada	1020.2	3.1%
6	South Korea	887.0	2.7%
7	India	873.2	2.7%
8	Australia	789.2	2.4%
9	Brazil	765.1	2.3%
10	Russia	509.7	1.6%

The European Union is the world's largest publisher in 2009, judging from the sample. It represents some 32% of world science, surpassing the 25% share of the next highest nation, the United States. Germany, United Kingdom, France, Italy, Spain and the Netherlands all report world publication shares in excess of the tenth entry in the table, the country Russia. Within the European Union, Germany has a small publication lead over other countries. China is the third highest publisher, with over 10% of world science. The People's Republic of China delivers over 87% of Chinese scientific output. China is the leader of the so-called "BRICK" nations, but Brazil, Russia, India and South Korean are all included on the table. Japan is the fourth highest nation. Two Commonwealth nations -- Canada and Australia – are also included in the leading publishers.

Nations in the sample markedly differ in their openness to collaboration with others (table 2). Of the countries listed, the Netherlands is the country most open to collaborating with other nations – nearly 16% of all its papers are written in collaboration with other nations. Other European nations show nearly as much openness. However when collaborative flows aggregated inside versus outside the European Union, it is clear that the European Union is not especially open to international research collaboration. Canada, Australia, Russia, the United States all exceed European Union levels of collaboration. Taiwan shows the lowest rates of international collaboration.

Table 2. Openness to Collaboration with Other Countries, Communities or Groups

Rank	Country or Community or Grouping	Total Publication	Total Publication within Unit	Openness
1	European Union	10494.1	9845.1	6.2%
	Germany	1926.9	1682.8	12.7%
	United Kingdom	1883.9	1636.1	13.2%
	France	1305.8	1129.7	13.5%
	Italy	1157.8	1026.8	11.3%
	Spain	982.6	883.8	10.0%
	Netherlands	571.3	482.8	15.5%
2	USA	8432.1	7861.7	6.8%

3	Chinas	3357.8	3201.0	4.7%
	Peoples R China	2842.5	2709.3	4.7%
	Taiwan	515.3	491.7	4.6%
4	Japan	2038.3	1934.5	5.1%
5	Canada	1020.2	883.6	13.4%
6	South Korea	887.0	840.5	5.2%
7	India	873.2	829.6	5.0%
8	Australia	789.2	701.4	11.1%
9	Brazil	765.1	727.5	4.9%
10	Russia	509.7	466.1	8.6%

Figure 1 shows two fractal distributions of publication and citation in the world sample. On the left is the publication productivity of countries. The countries are ranked by productivity from 1, the most productive, to 156, the least productivity. The logarithm of the rank is shown on the abscissa, the logarithm of the publication is shown on the ordinate. The equivalent law of productivity for individual scientists is Lotka's law [3]. Lotka's law is expected to show a strict power law with a coefficient of -2. Across countries, the distribution is more complex than Lotka's law implies.

There is a shallow distribution of productivity across the most productive countries, and a more steep distribution across the least productive countries. The distribution is either multi-fractal, with multiple governing processes, or there are issues associated with fractionating papers equally by authors. For instance, if a scientist has two organizational affiliations, one in a high productivity country, and one in a low productivity country, it might be appropriate to use prior knowledge to assign credit more consistently across nations.

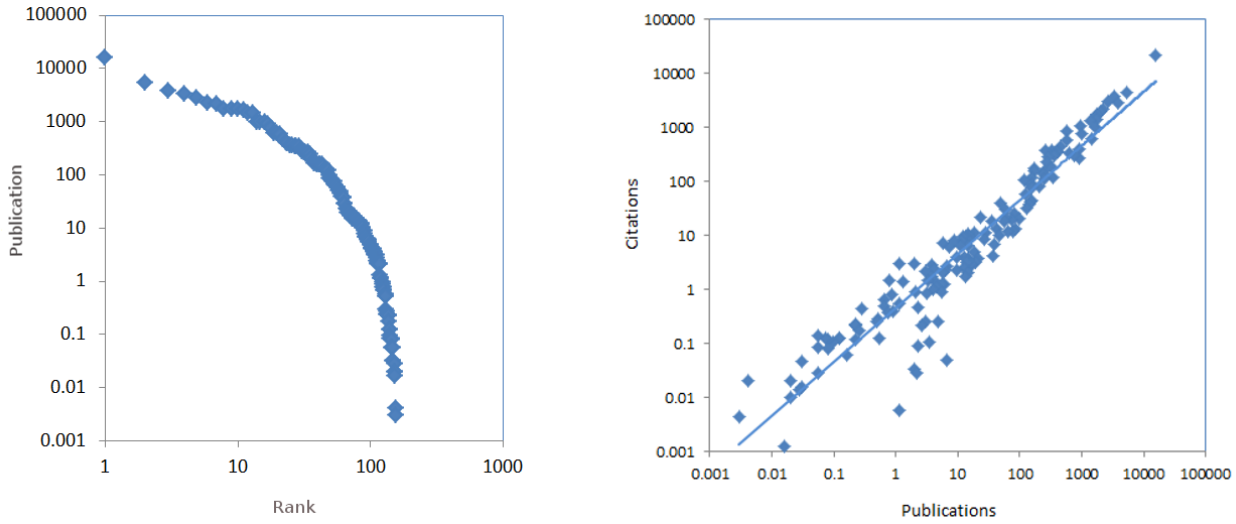

Figure 1: Fractal and Multi-Fractal Distributions in World Science

The right-most graph shows earned publications by earned citations across nations in the sample. Merton [4] noted an effect among individual scientists whereby the most published scientists earned the most citations. This "rich get richer" effect became known as the Matthew effect. Across nations, however, the Merton effect appears absent. There is a constant number of citations earned per paper. The coefficient on the power law is indistinguishable from 1.

5 Results from Edges

In section 4 we identified the major nodes of the world graph of science. The graph representation of science represents international collaboration as flows between the nodes. The total flow between major nodes in the graph is 2300.3 publications. Table 3 lists the top ten flows by magnitude, and also by fraction of the total. Nine of the ten flows are between the U.S. and other nations. The single largest flow is between the United States and the European Union, constituting 21% of the total flow. This flow is far in excess what would be expected if the flows were proportionate to raw publication. The magnitude of collaboration between the United States and Canada is also notable; this flow constitutes 5% of all international collaborations. U.S. collaborations with Israel are in the top ten largest flows, displacing potential collaboration with India, for instance. Finally, the flow of collaboration between Germany and Switzerland is particularly notable.

Table 3: Major Edges in the Graph of World Science

Rank	Nation	Nation	Total Publication	Percent of World International
1	USA	European Union	490.2	21.3%
2	USA	Canada	116.6	5.1%
3	USA	Peoples R China	102.6	4.5%
4	USA	Japan	59.3	2.6%

5	USA	Australia	44.9	2.0%
6	USA	South Korea	39.3	1.7%
7	USA	Switzerland	30.4	1.3%
8	Germany	Switzerland	28.8	1.3%
9	USA	Israel	24.4	1.1%
10	USA	Brazil	21.8	0.9%

It is helpful to also show substantial flows within the nodes themselves. These flows consist of three types: flows between member states, flows to the United States, and flows within the United Kingdom. Table 4 provides these statistics, including total publication, as well as the percent of world international collaboration. Note that these flows do not contribute to the count of world publication; this quantity is given only for comparative purposes.

Table 4: Major Flows Within Nodes

Nation	Nation	Total Publication	Percent of World International
European Union	European Union	741.7	32.2%
United Kingdom	Germany	45.0	2.0%
United Kingdom	France	30.7	1.3%
Germany	France	30.0	1.3%
Germany	Netherlands	26.1	1.1%
Germany	Italy	25.6	1.1%
France	Italy	23.4	1.0%
USA	European Union	490.2	21.3%
USA	United Kingdom	104.3	4.5%
USA	Germany	96.3	4.2%
USA	Italy	62.8	2.7%
USA	France	55.1	2.4%
USA	Spain	30.0	1.3%
USA	Netherlands	35.8	1.6%
United Kingdom	United Kingdom	47.6	2.1%
England	Scotland	34.0	1.5%

Both the United Kingdom and Germany play central roles within the European Union node. The United Kingdom has two of the highest magnitude links within the European Union node, and the single highest magnitude flow outside the European Union node. The pre-eminence of the United Kingdom occurs in part because of an aggregation of countries; as seen in the table there are major collaborative flows between England and Scotland. Germany also has extensive connections, with the United States, as well as within Europe. The United States shares strong research ties with all major European countries.

Table 5 provides the closeness centrality of the major nodes in the network. These statistics largely confirm what is reported by other network measures. The most central nodes in the network are the United States, and secondarily Europe (particularly as represented by England and Germany). There appear to be strong relationships between national scientific output and the centrality of that country in the world network. However the causal relationship is unclear. Does centrality promote greater knowledge production, perhaps because of a greater access to information and skilled resources? Or does greater knowledge production lead to a higher centrality as researchers, incidentally to the conduct of their research, invest in international collaboration? Regardless of the causation it appears that the United States is comparatively more central in the network than publications would suggest, while China is out publishing its otherwise somewhat peripheral location in the network.

Table 5. Network Centrality

Rank	Country or Community		Centrality
1	USA		0.781
2	European Union*		0.643
	France		0.713
	Germany		0.707
	United Kingdom*		0.679
		England	0.730
		Scotland	0.628
	Italy		0.659
	Spain		0.651
	Netherlands		0.648
	Czech Republic		0.638
	Austria		0.635
	Sweden		0.635
	Belgium		0.628
	Poland		0.618
	Greece		0.615
	Slovenia		0.599
	Denmark		0.596
	Finland		0.592
3	Canada		0.654
4	Switzerland		0.654
5	Japan		0.643

6	Australia		0.633
7	India		0.633
8	Bulgaria		0.613
9	Peoples R China		0.613
10	Russia		0.603

* Average values for these aggregated entries are reported.

The world collaboration network can be visualized on a world map. For each country, we used the latitude and longitude of the capital as location of the node. The strength of the edge is converted to gray scale using a logarithmic scaling. Only the edges that pas a threshold are visualized. The edges are drawn based on the great circle, using the shortest great circle for each connection. The nodes are scaled based on the cubic root of the magnitude of publications without international collaboration. Fig. 1 shows the resulting figure. This figure shows clearly that the US is very big in terms of internal production of scientific publications, but also how central it is in terms of collaborations. It appears that transatlantic links are dominant. Therefore, we zoomed in on this part of the map, resulting in 2. Note that here we visualize England, Scotland, Wales, and Ireland as separate nodes. Similarly, the individual member states of the EU are shown, instead of aggregating it. This figure shows not only the importance of transatlantic collaboration in science, it also reveals the presence of strong collaborative ties inside Europe. Fig. 3 shows this European network in more detail.

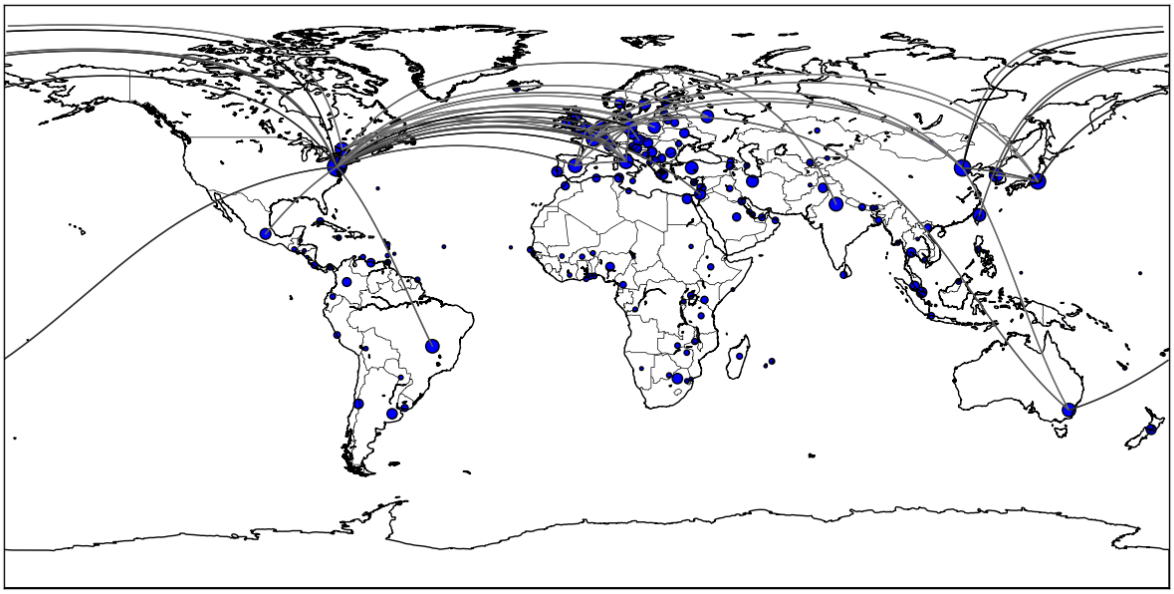

Figure 1: World Scientific collaboration network

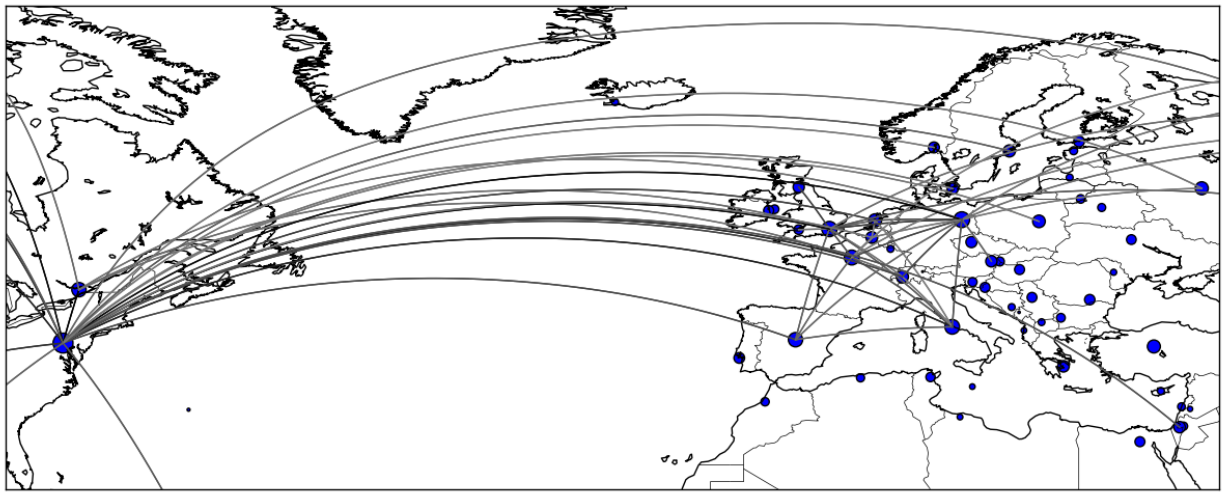

Figure 2: Transatlantic collaboration

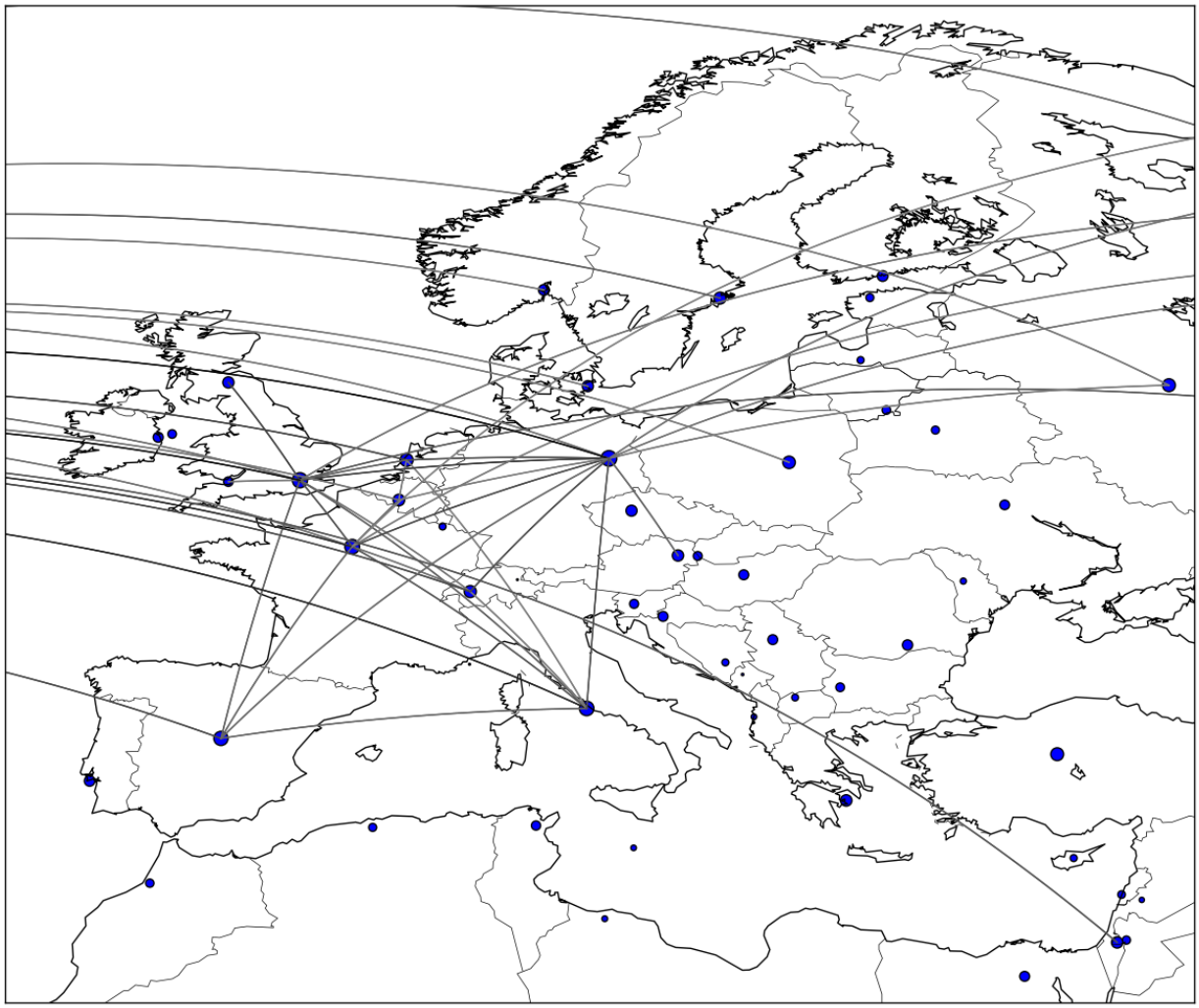

Figure 3: European Collaboration

In the following section we consider whether the morphology of the network is at least partially explicable through the use of a network regression technique. We hypothesize that the world science collaborative network is a random graph. That is, countries choose prospective partners randomly according to total volume of published output (table 1). There is however a certain degree of preferential attachment – countries prefer to work with geographically proximate partners, all things equal.

Testing these hypotheses requires two main variables. Distance is the logarithm of the great circle distance between the capitals of two countries. This is only a rough proxy for actual distance, but is useful as a first approximation. Publication is the logarithm of the geometric mean publication of the two countries. A completely random graph would show a parameter here of 0.500; connections are made in proportion to the total publication of the country, and selected in direct proportion to the publication output of other nations. The independent variable is the number of collaborative articles written between two countries in the sample, where collaboration has been evidenced in the sample. The resultant output of a regression model is then given below.

Table 6. Regression Model Explaining Collaboration Intensity

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.577	.332	.332	2.366

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13799.017	2	6899.508	1232.336	< 0.001
	Residual	27708.088	4949	5.599		
	Total	41507.105	4951			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	Constant	-3.844	.272		-14.146	< 0.001
	Distance	-.433	.033	-.155	-13.038	< 0.001
	Publication	.475	.010	.588	49.549	< 0.001

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
	Constant	-3.844	.272		-14.146	< 0.001
	Distance	-.433	.033	-.155	-13.038	< 0.001
	Publication	.475	.010	.588	49.549	< 0.001

The model shows a significant negative effect of distance between nations on the resulting collaboration. Prior publication is a significant predictor of publication. All variables in the model, and the model itself, can be rejected with a probability less than 0.001. The model shows significant departures from randomness, even once distance is taken into consideration. In particular, the publication parameter is significantly less than 0.500 (one-sided t-test, $p < 0.05$). This indicates that countries are disassortative in their selection of partners – countries are slightly less likely to take larger publishing nations as collaborative partners than would be expected by chance. Despite the effectiveness of this model, the collaboration between nations in world science remains deeply uncertain. Two thirds of the variance in the data remains unexplained.

6 Conclusions

This paper adopted a complex network perspective in surveying the world science system in 2009. The paper demonstrates the continued preeminence of the United States in the world network. European output is higher overall, however, with a strong bilateral and trilateral performance by England, Germany and France. The BRICK nations (Brazil, Russia, India, China and South Korea) firmly rank in the top publishing nations. China is arguably more peripheral in the world network than its publication output would otherwise suggestion. Whether this position is necessarily a bad thing depends on whether we regard network position as a cause, or a consequence, of scientific performance. The matter is currently unresolved.

Collaboration is especially high between the United States and Europe. Despite this, the United States maintains strong collaborative relationships with all major countries in the network. Asian collaborations between the U.S. and Europe, may soon add a third hub to the world network of science. A surprising finding is that nations vary widely in their openness to international collaboration. Two of the

most open nations are Canada and the Netherlands. Although the openness of these nations is unexplained, we may advance several hypotheses. First, it may be that bilingual policies permit a greater range of potential research partners. Secondly, it may be a matter of opportunity – both nations are large publishers which are neighbored by even larger publishing nations. The comparative advantages of partnership may be higher for these nations.

Several challenges to the random graph hypothesis were seen in the data. First, there was little or no scaling of citations by publication at the national level. Second, publishing output showed a complex scaling, suggesting either multiple regimes in the data, or a need for better fractionation of the data to take into account strong prior knowledge concerning world outputs of science. Third, the world scientific network can be partially explained in terms of random processes and preferential attachment. Despite this, the network is strongly shaped by geographical proximity. Further, the network is significantly disassortative. Larger publishing nations preferentially collaborate with smaller nations, especially if these nations are geographically proximate. There is only weak evidence that there are two regimes to the data; for instance that large publishers collaborate internationally, while small publishers are forced to seek partnerships at a local or continental scale. Nonetheless, this hypothesis should be further explored.

There are several avenues for further research. It would be useful to further consider the measurement of international collaboration. Apparently straightforward measures of collaboration, involving institutional affiliations and authors, are not entirely satisfactory in measuring international collaborative flows. A systematic exploration of the various techniques for fractionating papers, and thereby accounting for international collaboration should be undertaken. The measurement choices used appear to affect the resultant outcomes. A second avenue for investigation would be to consider the role of funding agencies. Some national and international funders of science are better able to enervate international collaboration. Measures and metrics of this phenomenon would be worth investigating. Further, a complex network perspective would be particularly useful; the network might be represented as a two mode graph for instance. Countries would be one mode of the graph, and funding agencies would be the other mode.

Knowledge on the graph is still poorly understood. Knowledge production is very heterogeneous across countries; it is not merely a matter of the West versus the rest. Collaborative links are similarly heterogeneous; it would be worthwhile to investigate this further. Current research has argued that organizational collaboration is strongly shaped by technological proximity [37]. Complementarities in knowledge assets might be a useful explanatory factor in understanding why certain nations collaborate so

intensely. The structure of the network, both at the node level as well as at the edge level, is still poorly understood. In particular, a two-staged model may be required. At the first stage, explanations for national output may be required. At the second stage, the aggregate decision to invest in international collaboration might be explained. Previous work has examined institutional as well as economic factors in explaining national output [38]. A similar effort might be extended to understanding publication and collaborative data. Openness to collaboration varies strongly in the sample, and often in an unexpected manner. For instance, Europe as a whole is relatively closed to foreign research activity.

Acknowledgements: The authors appreciate the research assistance provided by Stephen Carley at the Georgia Institute of Technology. All errors however are our own.

References

1. Zipf, G.K., *Human Behavior and the Principle of Least Effort*. 1949, Boston, MA: Addison-Wesley.
2. Bradford, S.C., *Sources of Information on Specific Subjects.* Engineering: An Illustrated Weekly Journal, 1934. **137**: p. 85-86.
3. Lotka, A.J., *The frequency distribution of scientific productivity.* Journal of the Washington Academy of Sciences, 1926. **16**(12): p. 317-324.
4. Merton, R.K., *The Matthew Effect in Science.* Science, 1968. **159**(3810): p. 56-63.
5. de Solla Price, D., *A General Theory of Bibliometric and Other Cumulative Advantage Processes.* Journal of the American Society for Information Science, 1976. **27**(5-6): p. 292-306.
6. Newman, M.E.J., *The structure of scientific collaboration networks.* Proceedings of the National Academy of Science, 2001. **98**: p. 404-409.
7. Godin, B., *Science, Accounting and Statistics: The Input-Output Framework.* Research Policy, 2007. **36** (9): p. 1388-1403.
8. Bush, V., *Science, the Endless Frontier*. 1945, Washington, D. C. : U. S. Government Printing Office.
9. Bush, V., *As We May Think*, in *The Atlantic*. 1945.
10. Freeman, C., *Measurement of Output of Research and Experimental Development,*. 1969, UNESCO.
11. de la Mothe, J., *The Revision of International Science Indicators: The Frascati Manual.* Technology in Society, 1992. **14**: p. 427-440.
12. Bourke, P. and L. Butler, *Standards issues in a national bibliometric database: The Australian case.* Scientometrics, 1996. **35**(2): p. 199-207.
13. Bureau of Industry Economics, *Australian Science: Performance from Published Papers*. 1996, Australian Government Publishing Service: Canberra.
14. Shapira, P., et al., *Knowledge Economy Measurement: Methods, Results and Insights from the Malaysian Knowledge Content Study.* Research Policy, 2006. **35**: p. 1522-1537.
15. Ministerie van Onderwijs Cultuur en Wetenschap. *Wetenschap en Technologie-Indicatoren* 2010 March 4, 2011]; Available from: http://www.nowt.nl/docs/NOWT-WTI_2010.pdf.
16. Katz, J.S. and D.M. Hicks, *A Systemic View of British Science.* Scientometrics, 1996. **15**(1): p. 133-154.
17. National Science Board. *Science and Engineering Indicators*. 2010 March 4, 2011]; NSB 10-01:[Available from: http://www.nsf.gov/statistics/seind10/pdf/seind10.pdf.

18. Hicks, D.M., *Systemic data infrastructure for innovation policy*, in *Science of Science Measurement Workshop*. 2010, NSTC's Interagency Task Group: Washington, D. C. .
19. Ministerio da Ciencia e Tecnologia. *Plataforma Lattes*. 2011; Available from: http://lattes.cnpq.br/english/index.htm.
20. European Commission, ed. *Towards a European Research Area Science, Technology and Innovation - Key Figures*. 2003, European Commission Brussels.
21. Porter, A.L. and J.D. Roessner, *Indicators of national competitiveness in high technology industries*. 1991, Science Indicators Studies Group, National Science Foundation: Washington, DC.
22. Porter, A.L., et al., *Indicators of high technology competitiveness of 28 countries.* International Journal of Technology Management, 1996. **12**(1): p. 1-32.
23. Katz, J.S. and B.R. Martin, *What is research collaboration?* Research Policy, 1995. **26**(1): p. 1-18.
24. Esler, S.L. and M.L. Nelson, *Evolution of scientific and technical information distribution.* Journal of the American Society for Information Science, 1998. **49**(1): p. 82-91.
25. Bozeman, B., *Technology transfer and public policy: A review of research and theory.* Research Policy, 2000. **29**(4-5): p. 627-655.
26. Dasgupta, D. and P.A. David, *Towards a new economics of science.* Research Policy, 1994. **23**(5): p. 487-52`.
27. Lewison, J.G.G., *Government Funding of Research and Development.* Science, 1997. **278**(5339): p. 878-880.
28. Buisseret, T.J., H.M. Cameron, and L. Georghiou, *What Difference Does it Make? Additionality in the Public Support of R&D in Large Firms.* International Journal of Technology Management, 1995. **10**(4-6): p. 587-600.
29. Wagner, C.S. and L. Leydesdorff, *Network structure, self-organization, and the growth of international collaboration.* Research Policy, 2005. **34**(10): p. 1608-1618.
30. Leydesdorff, L. and I. Rafols, *A Global Map of Science Based on the ISI Subject Categories.* Journal of the American Society for Information Science and Technology, 2009. **60**(2): p. 348-362.
31. Boyack, K.W., R. Klavans, and K. Börner, *Mapping the backbone of science.* Scientometrics, 2005. **64**(3): p. 351-374
32. Börner, K., C. Chen, and K. Boyack, *Visualizing Knowledge Domains.* Annual Review of Information Science & Technology in Society, 2003. **37**(1): p. 179-255.
33. Small, H., *Tracking and predicting growth areas in science.* Scientometrics, 2006. **68**(3): p. 595-610.
34. Small, H., *Co-citation context analysis and the structure of paradigms.* Journal of documentation, 1980. **36**: p. 183-196.
35. van Raan, A.F.J., *Fractal Dimension of Cocitations.* Nature, 1990. **347**(6294): p. 626.
36. Porter, A.L. and S.W. Cunningham, *Tech Mining: Exploiting New Technologies for Competitive Advantage*. 2005, New York: John Wiley & Sons.
37. Cunningham, S.W. and C. Werker, *Proximity and Collaboration in Regional Science.* Papers in Regional Science, under submission.
38. Cunningham, S.W. *A Comparative Political Theory of National Science Provision*. in *Atlanta Conference on Science and Innovation Policy*. 2009. Atlanta, GA: IEEE.